

A feed-forward neural network as a nonlinear dynamics integrator for supercontinuum generation

Lauri Salmela^{1,*}, Mathilde Hary^{1,2}, Mehdi Mabed², Alessandro Foi³, John M. Dudley², and Goëry Genty¹

¹Photonics Laboratory, Physics Unit, Tampere University, 33014 Tampere, Finland

²Institut FEMTO-ST, Université Bourgogne Franche-Comté CNRS UMR 6174, 25000 Besançon, France

³Laboratory of Signal Processing, Tampere University, 33014 Tampere, Finland

*Corresponding author: lauri.salmela@tuni.fi

November 23, 2021

Abstract

The nonlinear propagation of ultrashort pulses in optical fiber depends sensitively on both input pulse and fiber parameters. As a result, optimizing propagation for specific applications generally requires time-consuming simulations based on sequential integration of the generalized nonlinear Schrödinger equation (GNLSE). Here, we train a feed-forward neural network to learn the differential propagation dynamics of the GNLSE, allowing emulation of direct numerical integration of fiber propagation, and particularly the highly complex case of supercontinuum generation. Comparison with a recurrent neural network shows that the feed-forward approach yields faster training and computation, and reduced memory requirements. The approach is generic and can be extended to other physical systems.

Neural networks (NNs) are a central subset of machine learning techniques widely used in data analysis, classification and prediction [1]. A central aspect of NNs is their ability to link the input and output of a multidimensional system, of particular benefit for modeling complex and analytically intractable relationships as is typically the case with nonlinear systems. Indeed, several studies have demonstrated the use of NNs to forecast nonlinear evolution using including physics-informed methods [2–4], data-driven approaches [5–7], and hybrid techniques [8, 9]. In optics, NNs are becoming increasingly applied to study multidimensional ultrafast and chaotic systems [10], with recent applications including the optimization of mode-locked lasers [11–13] and the analysis of ultrafast instabilities [14–16].

A particular focus has been the use of NNs to study nonlinear propagation and supercontinuum (SC) generation in optical fibre [14, 15, 17], a complex process involving multiple nonlinear and dispersive effects [18]. Both the propagation dynamics and the output spectral and temporal characteristics depend sensitively on the injected pulse

and fiber parameters, and matching input conditions to achieve a desired output is a complex multivariate problem. The traditional approach for optimization is based on parameter scanning using step-by-step integration of the generalized nonlinear Schrödinger equation (GNLSE) [19]. Yet whilst the GNLSE has been shown to accurately model fiber nonlinear dynamics, direct simulations are time consuming, especially with a large parameter space of potential boundary conditions.

To overcome this limitation, attempts have been made to use machine learning techniques to optimize and control fiber dynamics, with one approach being the use of genetic algorithms to tailor broadband SC spectra [20, 21]. More recently, recurrent neural networks (RNNs) using only the input temporal (or spectral) intensity profile of the injected pulse have been shown to emulate fiber propagation dynamics [15] with accurate prediction of SC evolution maps in computation times as short as one second. A limitation, however, is that this approach requires an initial training phase of several hours due to the multiple iterative loops associated with the RNN internal memory.

Here, we show how the full-field (intensity and phase) evolution of ultrashort pulses in optical fiber can be accurately modeled with a faster and simpler feed-forward neural network (FNN) over a wide range of input pulse properties (peak power, duration, chirp), and fiber parameters (dispersion, nonlinearity). The key conceptual novelty is that we train the network to learn the differential propagation dynamics of the GNLSE i.e. to accurately replicate the change in intensity and phase of the electric field between elementary steps. Once trained on the differential dynamics, the network then can model the long-term evolution from a given input. We also perform a detailed comparison with a RNN model, highlighting the benefits of the FNN approach in terms of speed and memory.

The principle is illustrated in Fig. 1. We first generate an ensemble of pulse propagation data corresponding to highly complex broadband coherent SC generation. The dynamical map is completely characterized by a vector $[I_n(z_i, X), \Phi_n(z_i, X)]$, where I_n and Φ_n represent intensity and phase at some distance z_i , expressible either in the temporal ($X = T$) or spectral domains ($X = \omega$), and from which the complex electric field can be reconstructed. The subscript $n = 1 \dots N$ indicates a particular map for a distinct set of input pulse and fiber parameters.

This data is generated by numerically integrating the GNLSE with the split-step method, seeded by hyperbolic-secant input pulses at $\lambda_0 = 830$ nm, with peak power and duration (FWHM) in the range $P_0 = 0.77\text{--}1.43$ kW and $T_{\text{FMHM}} = 70\text{--}130$ fs ($\pm 30\%$ variation). The fiber dispersion parameters are: $\beta_2 = -5.90 \times 10^{-27}$ s²m⁻¹, $\beta_3 = 4.21 \times 10^{-41}$ s³m⁻¹, $\beta_4 = -1.25 \times 10^{-55}$ s⁴m⁻¹, and $\beta_5 = -2.45 \times 10^{-70}$ s⁵m⁻¹ (zero-dispersion wavelength at 767 nm), and the nonlinear coefficient is: $\gamma = 0.1\text{W}^{-1}\text{m}^{-1}$. The fiber length is $L = 20$ cm.

The key idea shown in Fig. 1(a) is to teach the network the differential change in intensity and phase associated with an elementary propagation distance Δz . To achieve a performance advantage relative to direct integration, the aim is to use a significantly larger step in the FNN compared to that used in GNLSE integration. To this end, the intensity and phase evolution are downsampled at distances $z_i = (i-1)\Delta z$ ($i = 1..M$), where $\Delta z = L/M = 0.1$ cm is 50 times larger than in the GNLSE simulations used to generate the data. The downsampled vectors are then used as the

FNN input. The network output vectors after an elementary step Δz are $[I_n(z_{i+1}, X), \Phi_n(z_{i+1}, X)]$. The change in the intensity and phase modeled by the FNN is then compared to that from the GNLSE via an error function [14].

Once trained, the neural network acts as a very fast and memory-efficient GNLSE integrator. It can predict the intensity and phase $[I(z + \Delta z, X), \Phi(z + \Delta z, X)]$ after an elementary propagation distance Δz given the complex field $[I(z, X), \Phi(z, X)]$ at distance z , from which the dynamical evolution of the complex electric field can be reconstructed. The trained FNN can then be used to predict propagation dynamics over an extended distance using an iterative loop (see Fig. 1(b)) such that the intensity and phase $[I(z_{i+1}, X), \Phi(z_{i+1}, X)]$ are fed back to the network as a new input to predict the field amplitude $[I(z_{i+2}, X), \Phi(z_{i+2}, X)]$ at distance $z + 2\Delta z$. This operation is performed over the full propagation distance.

The neural network itself consists of 3 hidden layers of 2000 nodes with ReLU activation ($f(x) = \max(0, x)$) and a sigmoid output layer with 2048 nodes. The codes were written in Python using Keras with Tensorflow backend [22]. The network is trained for 80 epochs with RMSprop optimizer and adaptive learning rate. The network can be trained in the temporal or spectral domain and with data input on either linear or logarithmic (dB) scales. In the results shown below we used ensembles of spectral evolution maps in logarithmic scale. Examples of time domain evolution using linear input are in the Supplementary information (Fig. S1). The accuracy of the network is tested with a separate set of propagation maps not used in the training phase. We quantify performance using the average (normalized) root mean squared (RMS) error:

$$R = \sqrt{\frac{\sum_{d,i} (x_{n,d,i} - \hat{x}_{n,d,i})^2}{\sum_{d,i} (x_{n,d,i})^2}}, \quad (1)$$

where \mathbf{x}_n and $\hat{\mathbf{x}}_n$ denote GNLSE simulation and FNN prediction for a particular realization n . Variables d and i indicate summation over intensity (spectral or temporal) and propagation steps, respectively. When evaluating performance over an ensemble, the error is calculated over N distinct evolution maps.

We first show the ability of the FNN to predict SC evolution developing from transform-limited input pulses. Here, we used an ensemble of 1500 simulations: 1400 for training phase and 100 for

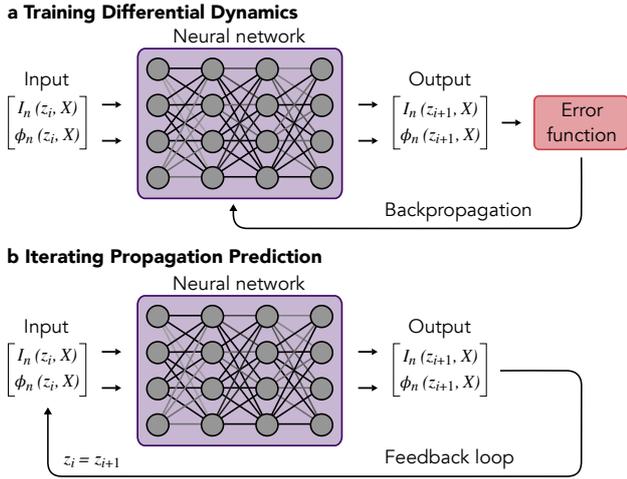


Figure 1: Feed-forward neural nonlinear dynamics integrator principle. **a** Training Differential Dynamics. Training is done from multiple input/output pairs generated by direct integration of the GNLSE and corresponding to the temporal ($X = T$) or spectral ($X = \omega$) intensity I and phase Φ of the propagating field at distances separated by an elementary step Δz (see text for details). The network variables are adjusted via gradient descent backpropagation. **b** Iterating Propagation Prediction. Once trained, the network effectively acts a GNLSE integrator and predict iteratively the intensity and phase evolution via feedback loop. The prediction is initialized from the intensity and phase profile at the fiber input.

testing. Predicted spectral evolution maps are shown in Fig. 2 for input peak power and pulse duration of 1.32 kW and 120 fs (Fig. 2a). For comparison, we also plot the evolution from direct GNLSE integration. The RMS error for the realizations shown in Fig. 2a is $R = 0.098$, while the average error computed over the 100 test evolution maps is $R = 0.094$. The neural network accurately predicts the SC development, with dispersive wave and soliton dynamics reproduced over a ~ 40 dB dynamic range.

We next tested the ability of the network to model SC development from chirped pulses. Here, we performed 3000 simulations with the same parameters as above, except with peak power variation of $\pm 20\%$ and input pulse spectral bandwidth varying from transform-limited (TL) to twice the TL with random sign of chirp. The SC spectral evolution predicted by the network for pulses with 942 W peak power, 84 fs duration, and positive chirp of 1.53 times the TL bandwidth are shown Fig. 2b. Again in this case we see how the main features including the spectral interference

fine structure are well-reproduced ($R = 0.190$) by the FNN model although we do note a small discrepancy in the distance of maximum compression at -20 dB bandwidth. The RMS error $R = 0.383$ (0.242 median) computed over the 100 test ensemble shows that the network also accurately models chirped pulse dynamics.

The results above correspond to the case of anomalous dispersion regime SC generation, but the network can be trained over a much wider range of dynamics. To this end, one can use the normalized form of the GNLSE to generate the training ensemble of evolution maps (see Supplementary information), and map dimensional parameters to normalized values to predict the evolution corresponding to a specific set of parameters. For example, Fig. 3 plots examples of predicted SC evolution for a pump wavelength in the normal dispersion region (see caption for parameters). Specifically, Fig. 3a shows results for a TL limited pulse injected near the zero-dispersion wavelength while Fig. 3b shows the spectral evolution for a pump wavelength further detuned into the normal dispersion regime. We observe very good accuracy with $R = 0.141$ for Fig. 3a and $R = 0.043$ for Fig. 3b. The RMS error over an ensemble of 200 realizations is $R = 0.060$. Predictions in the time domain can be found in the Supplementary information (Fig. S2).

To reduce computational memory and increase the speed in the training phase, one can train the network from convolved spectral intensity and phase evolution maps. At first sight, a disadvantage of using convolved data is that the resulting wavelength/frequency grid is no longer on a Fourier grid, requiring separate training to predict spectral and temporal evolution. However, this is in fact a major benefit, because it allows to appropriately select the resolution in the spectral or temporal domains to optimally capture the relevant physical structure.

Results of predicted spectral evolution maps using convolved spectral intensity and phase training data with a 8 nm FWHM super-Gaussian spectral filter are shown in Fig. 4a,b. These results correspond to the same input pulse and fiber parameters as in Fig. 2a,b. We see how the network predictions remain accurate with a mean convolved (logarithmic) spectral intensity RMS of 0.06 and 0.16 calculated over 100 distinct test evolution maps for the transform-limited and chirped cases, respectively. Other examples of predictions using different spectral resolution can be found in the

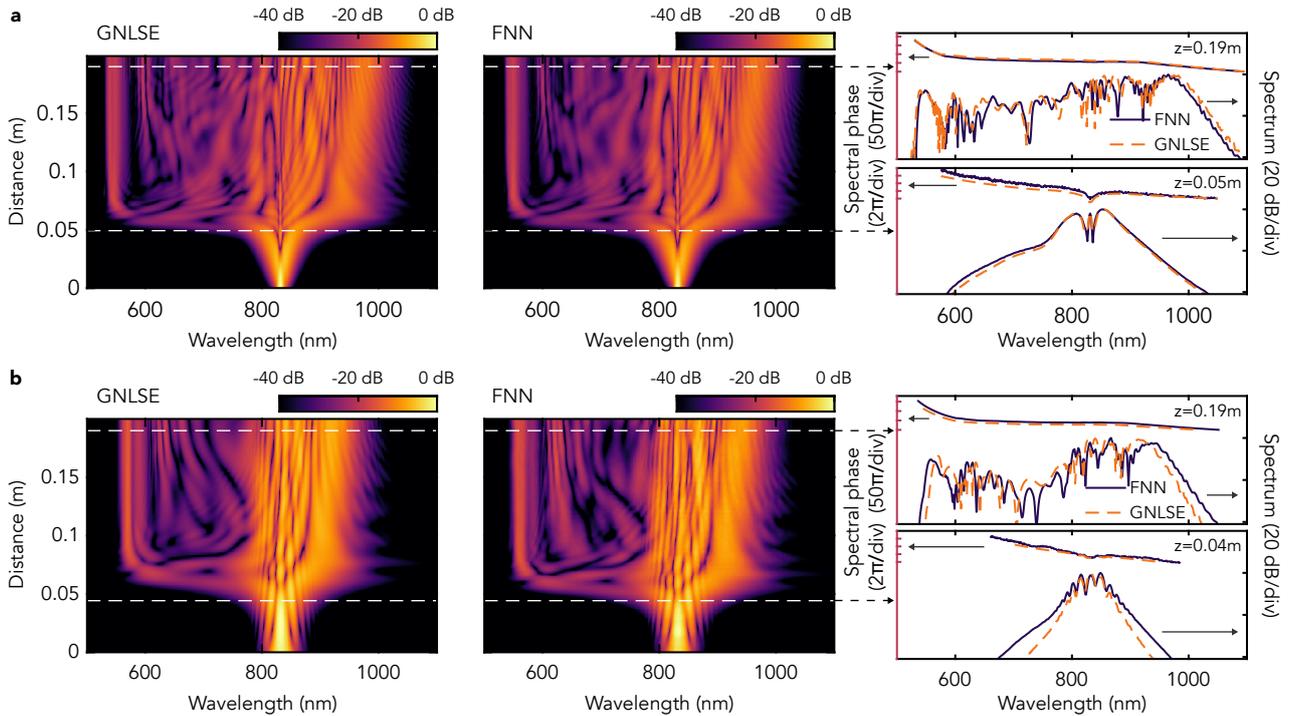


Figure 2: Comparison of supercontinuum spectral intensity evolution between GNLSE simulations (left panel) and neural network prediction (FNN, middle panel). The right panel shows the spectral intensity and phase at selected distances as indicated by the arrows. **a** shows the spectral intensity evolution of transform limited input pulses for input peak power and pulse duration values of $P_0 = 1.32\text{kW}$ and $T_{FWHM} = 120\text{fs}$, respectively. **b** shows the spectral intensity evolution for a chirped input pulse for input peak power and pulse duration values of $P_0 = 942\text{W}$ and $T_{FWHM} = 84\text{fs}$ with initial positive chirp of 1.53 times the TL bandwidth.

Supplementary information Fig. S3.

We then conducted a full comparison test of the computation resources and performance between the FNN model and an RNN similar to that used in Ref. [15]. The comparison was performed over an ensemble of 12,000 (11,800 for training and 200 for testing) convolved SC evolution maps in the anomalous dispersion regime with variations in peak power, pulse duration and dispersion (see Supplementary information). Table 1 summarizes the results, with examples of predicted maps shown in the Supplementary information (Fig. S4). For completeness we also list the computational resources used by the GNLSE simulations. Both FNN and RNN used the same number of free parameters/network variables, but the RNN is trained from spectral intensity maps which reduces by half the number of grid points compared to the FNN that includes both intensity and phase. The computational advantage of the FNN is clear. Specifically, FNN training and simulation times are reduced by a factor of four and five respectively, while memory usage during training is decreased by a factor of two. As might

	GNLSE	RNN	FNN
RMS error	N/A	R = 0.09	R = 0.19
Training time*	N/A	7.7 h	1.9 h
Simulation time**	38 min	1.6 s	0.35 s
Memory*	79 GB	7.7 GB	3.2 GB
Network var.	N/A	600k	600k
Num. points	8,192	132	264

Table 1: Comparison between normalized GNLSE numerical simulations, recurrent neural network (RNN) [15], and feed-forward neural network (RNN) for convolved spectral data. *11800 sims **200 sims

be expected from the faster computation, the FNN does show increased error compared to the RNN, but this does not lead to any significant visual differences in the evolution maps seen in the figures.

These results have shown model-free prediction of the full-field dynamics of ultrashort pulse propagation in optical fiber based on a feed-forward neural network trained to recognize differential propa-

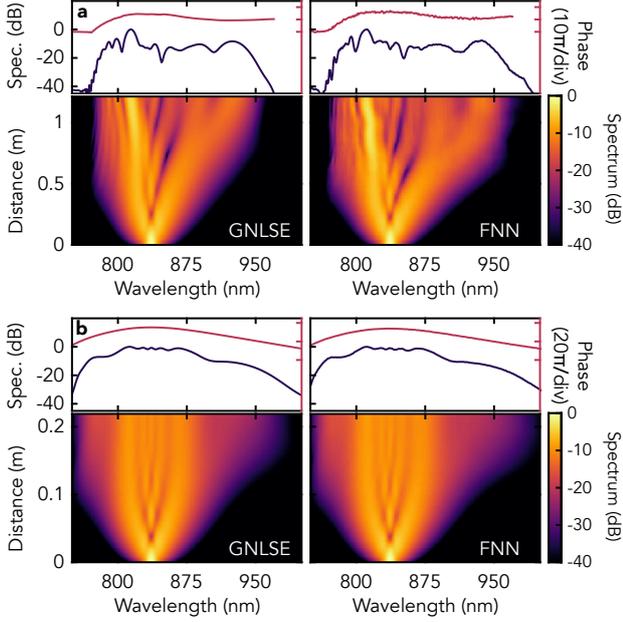


Figure 3: Spectral intensity evolution from simulations (GNLSE, left panel) and predicted by the neural network (FNN, right panel) for normal near-zero-dispersion pumping in **a** ($\gamma = 0.01 \text{ W}^{-1}\text{m}^{-1}$, $\beta_2 = 1.3 \times 10^{-27} \text{ s}^2\text{m}^{-1}$, $\beta_3 = 2 \times 10^{-41} \text{ s}^3\text{m}^{-1}$, $P_0 = 2.0 \text{ kW}$, $\lambda_0 = 835 \text{ nm}$, $T_{\text{FWHM}} = 100 \text{ fs}$) and far-normal pumping in **b** ($\gamma = 0.01 \text{ W}^{-1}\text{m}^{-1}$, $\beta_2 = 7.2 \times 10^{-27} \text{ s}^2\text{m}^{-1}$, $\beta_3 = 2 \times 10^{-41} \text{ s}^3\text{m}^{-1}$, $P_0 = 13.6 \text{ kW}$, $\lambda_0 = 835 \text{ nm}$, $T_{\text{FWHM}} = 100 \text{ fs}$). The top panels show the spectral intensity and phase at the fiber output.

gation dynamics within a GNLSE model. As compared to the recently introduced RNN approach, this FNN method is simpler and possesses significant advantages in terms of speed and memory. We expect our results to be of significance for the real-time optimization and control of nonlinear dynamics and we anticipate this approach could become a standard tool in nonlinear physics.

Funding. Faculty of Engineering and Natural Sciences graduate school of Tampere University. French Agence Nationale de la Recherche (ANR-15-IDEX-0003, ANR-17-EURE-0002). Academy of Finland (298463, 318082, Flagship PREIN 320165).

Disclosures. The authors declare no conflicts of interest.

Data availability. The data in this paper may be obtained from the authors upon request.

See Supplement 1 for supporting content.

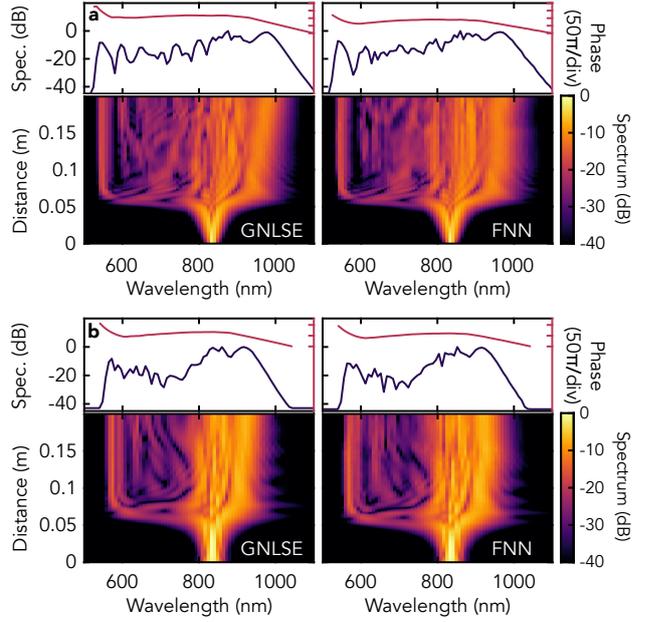


Figure 4: Comparison between the spectral evolution from simulations (GNLSE, left panel) and predicted by the neural network (FNN, right panel) when using convolved evolution maps for training with parameters identical to those in Fig. 2a and b. The top panels show the spectral intensity and phase at the fiber output.

References

- [1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, 2016.
- [2] S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Discovering governing equations from data by sparse identification of nonlinear dynamical systems,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.
- [3] M. Raissi, “Deep hidden physics models: Deep learning of nonlinear partial differential equations,” *Journal of Machine Learning Research*, vol. 19, no. 1, pp. 932–955, 2018.
- [4] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.
- [5] P. R. Vlachas, J. Pathak, B. R. Hunt, T. P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos, “Backpropagation algorithms and reservoir computing in recurrent

- neural networks for the forecasting of complex spatiotemporal dynamics,” *Neural Networks*, 2020.
- [6] J. Jiang and Y.-C. Lai, “Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: Role of network spectral radius,” *Physical Review Research*, vol. 1, no. 3, p. 033056, 2019.
- [7] M. Sangiorgio and F. Dercole, “Robustness of LSTM neural networks for multi-step forecasting of chaotic time series,” *Chaos, Solitons & Fractals*, vol. 139, p. 110045, 2020.
- [8] A. Wikner, J. Pathak, B. Hunt, M. Girvan, T. Arcomano, I. Szunyogh, A. Pomerance, and E. Ott, “Combining machine learning with knowledge-based modeling for scalable forecasting and subgrid-scale closure of large, complex, spatiotemporal systems,” *Chaos*, vol. 30, no. 5, p. 053111, 2020.
- [9] Y. Lei, J. Hu, and J. Ding, “A hybrid model based on deep LSTM for predicting high-dimensional chaotic systems,” *arXiv preprint arXiv:2002.00799*, 2020.
- [10] G. Genty, L. Salmela, J. M. Dudley, D. Brunner, A. Kokhanovskiy, S. Kobtsev, and S. K. Turitsyn, “Machine learning and applications in ultrafast photonics,” *Nature Photonics*, vol. 15, pp. 91–101, 2020.
- [11] G. Pu, L. Yi, L. Zhang, and W. Hu, “Intelligent programmable mode-locked fiber laser with a human-like algorithm,” *Optica*, vol. 6, no. 3, pp. 362–369, 2019.
- [12] U. Andral, R. S. Fodil, F. Amrani, F. Billard, E. Hertz, and P. Grellu, “Fiber laser mode locked through an evolutionary algorithm,” *Optica*, vol. 2, no. 4, pp. 275–278, 2015.
- [13] A. Kokhanovskiy, A. Ivanenko, S. Kobtsev, S. Smirnov, and S. Turitsyn, “Machine learning methods for control of fibre lasers with double gain nonlinear loop mirror,” *Scientific Reports*, vol. 9, no. 1, p. 2916, 2019.
- [14] M. Närhi, L. Salmela, J. Toivonen, C. Billet, J. M. Dudley, and G. Genty, “Machine learning analysis of extreme events in optical fibre modulation instability,” *Nature Communications*, vol. 9, no. 1, pp. 1–11, 2018.
- [15] L. Salmela, N. Tsipinakis, A. Foi, C. Billet, J. M. Dudley, and G. Genty, “Predicting ultrafast nonlinear dynamics in fibre optics with a recurrent neural network,” *Nature Machine Intelligence*, vol. 3, no. 4, pp. 344–354, 2021.
- [16] P. Amil, M. C. Soriano, and C. Masoller, “Machine learning algorithms for predicting the amplitude of chaotic laser pulses,” *Chaos*, vol. 29, no. 11, p. 113111, 2019.
- [17] L. Salmela, C. Lapre, J. M. Dudley, and G. Genty, “Machine learning analysis of rogue solitons in supercontinuum generation,” *Scientific Reports*, vol. 10, no. 1, pp. 1–8, 2020.
- [18] J. M. Dudley, G. Genty, and S. Coen, “Supercontinuum generation in photonic crystal fiber,” *Reviews of Modern Physics*, vol. 78, no. 4, p. 1135, 2006.
- [19] G. Agrawal, *Nonlinear fiber optics*. Academic Press, 5th ed., 2013.
- [20] B. Wetzell, M. Kues, P. Roztocki, C. Reimer, P.-L. Godin, M. Rowley, B. E. Little, S. T. Chu, E. A. Viktorov, D. J. Moss, A. Pasquazi, M. Peccianti, and R. Morandotti, “Customizing supercontinuum generation via on-chip adaptive temporal pulse-splitting,” *Nature Communications*, vol. 9, no. 1, pp. 1–10, 2018.
- [21] L. Michaeli and A. Bahabad, “Genetic algorithm driven spectral shaping of supercontinuum radiation in a photonic crystal fiber,” *Journal of Optics*, vol. 20, no. 5, p. 055501, 2018.
- [22] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, “Tensorflow: A system for large-scale machine learning,” in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pp. 265–283, 2016.

Supplementary information

Supplementary results

Same evolution scenarios as in Figure 2 of the main manuscript but with training and prediction performed in the time domain. The RMS errors for temporal intensity are $R = 0.54$ and $R = 0.70$ for the transform-limited and chirped supercontinuum cases, respectively.

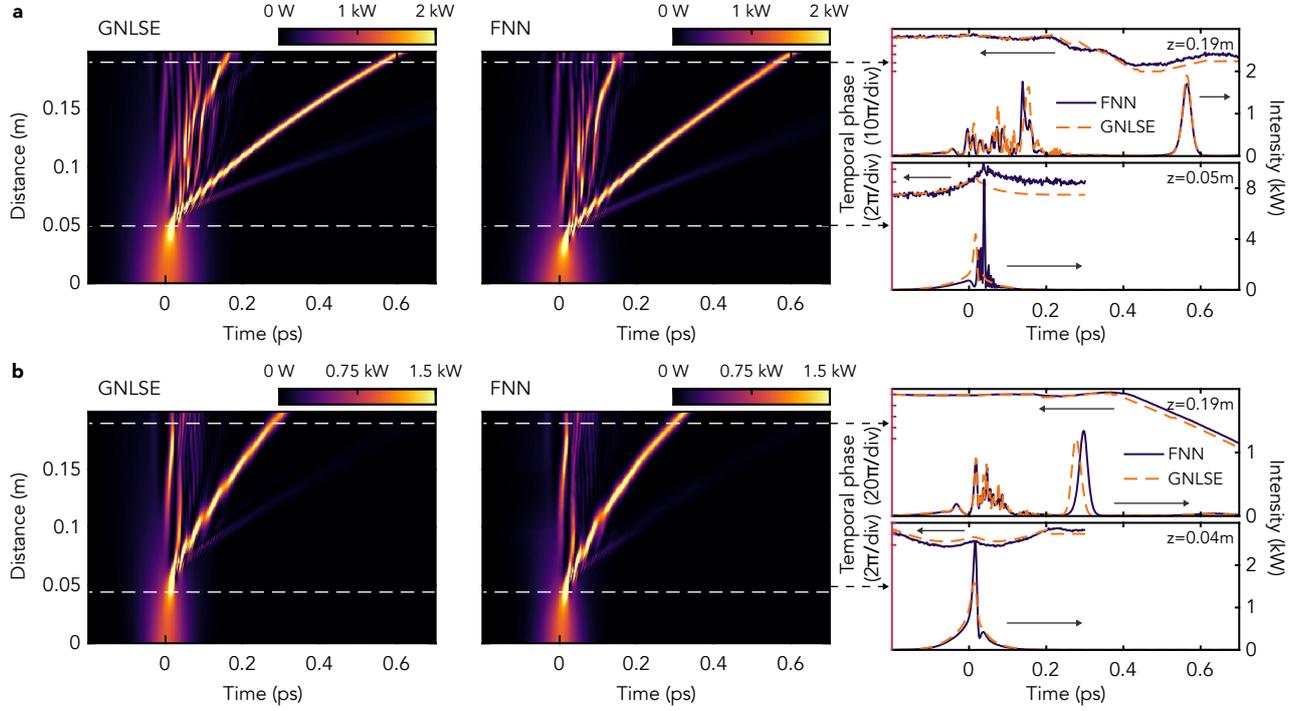


Figure S1: Temporal intensity evolution of supercontinuum from simulations (GNLSE, left panel), predicted by the neural network (FNN, middle panel), and their comparison at selected distances (right panel). **a** shows the evolution of a transform limited input pulse, and **b** shows the evolution for a chirped input pulse (see main text for details).

Same evolution scenarios as in Figure 3 of the main manuscript but with training and prediction performed in the time domain. The RMS error is $R = 0.17$ over 200 test evolution maps.

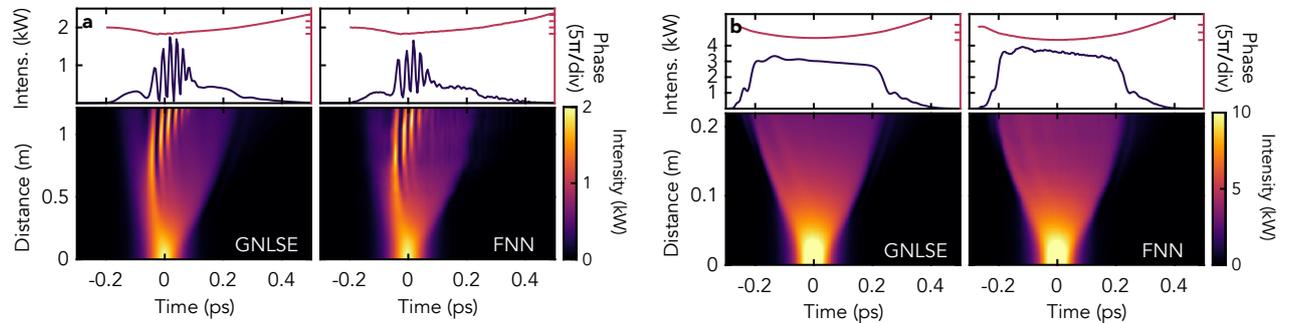


Figure S2: Temporal intensity evolution of supercontinuum from simulations (GNLSE, left panel), predicted by the neural network (FNN, right panel) for normal near-zero-dispersion pumping in **a** and far-normal pumping in **b**. The top panels show the temporal intensity and phase at the fiber output

Same evolution scenarios as in Figure 4a,b but with training intensity and phase data convolved with a 4 nm FWHM super-Gaussian filter. The (logarithmic) RMS error with the

4 nm resolution is $R = 0.06$ and $R = 0.17$ for the transform-limited and chirped cases, respectively, similar values as observed with the 8 nm resolution.

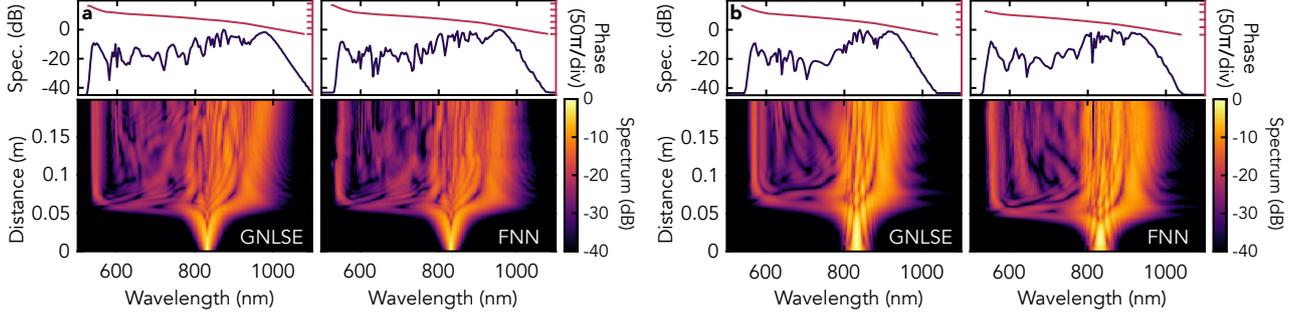


Figure S3: Convolved spectral intensity evolution of supercontinuum from simulations (GNLSSE, left panel), predicted by the neural network (FNN, right panel) and corresponding to the cases illustrated in Fig. 4a and b. The top panels show the spectral intensity and phase at the fiber output.

Comparison between recurrent and feed-forward neural network The (logarithmic) RMS errors over 200 test realizations are $R = 0.09$ and $R = 0.19$ for the recurrent and feed-forward neural networks, respectively.

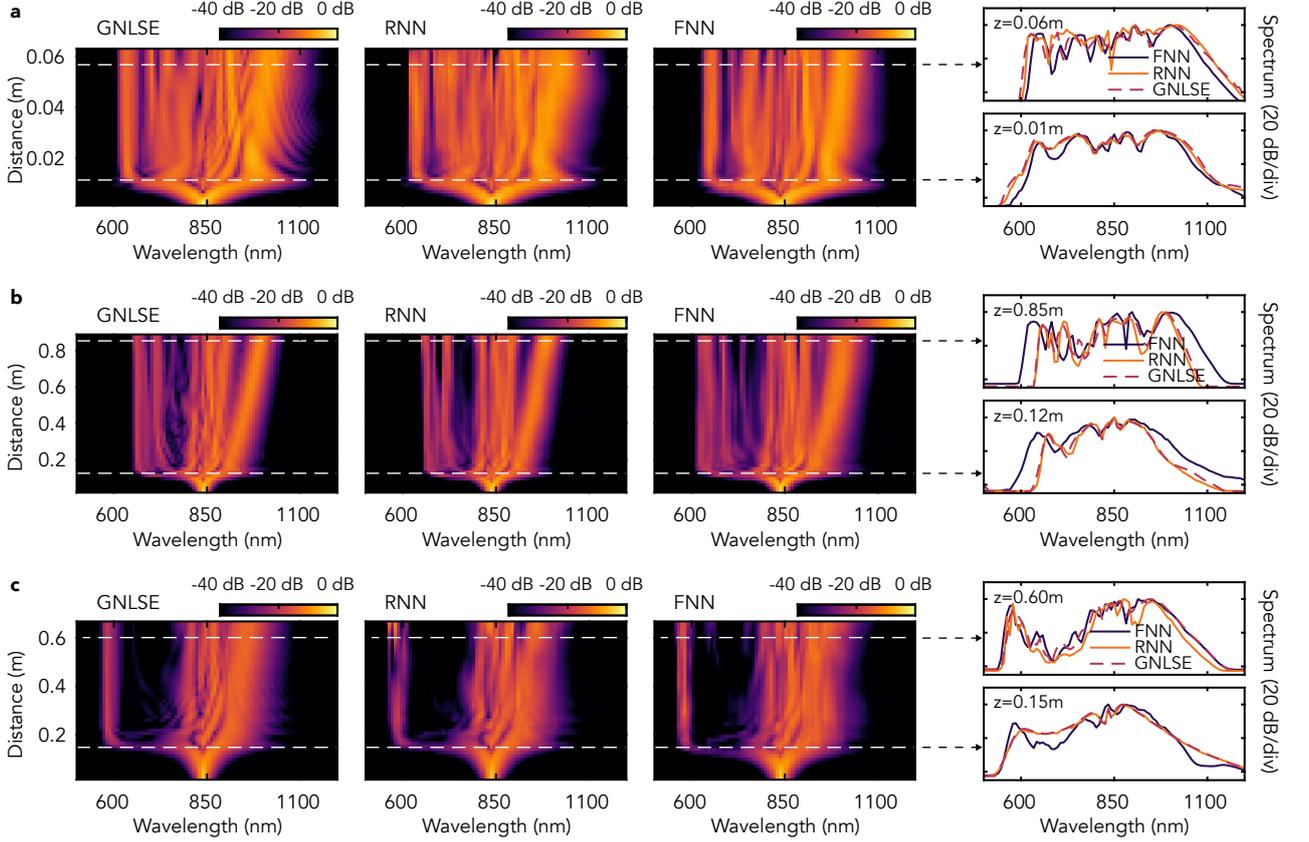


Figure S4: Convolved spectral intensity evolution of supercontinuum from simulations (GNLSSE), predicted by the recurrent (RNN) and neural network (FNN), and their comparison at selected distances (right panel). **a** shows the predictions for sech-type input pulse centered at 830 nm with 7.6 kW peak power and 40 fs duration with fibre parameters of $\gamma = 0.1 \text{ W}^{-1}\text{m}^{-1}$, $\beta_2 = -8 \times 10^{-27} \text{ s}^2\text{m}^{-1}$ and $\beta_3 = 9 \times 10^{-41} \text{ s}^3\text{m}^{-1}$. **b** shows the results for an input pulse with 2.9 kW peak power and 120 fs duration, and $\gamma = 0.0184 \text{ W}^{-1}\text{m}^{-1}$, $\beta_2 = -5.1 \times 10^{-27} \text{ s}^2\text{m}^{-1}$ and $\beta_3 = 4.3 \times 10^{-41} \text{ s}^3\text{m}^{-1}$. **c** shows the results for a realization with 3.0 kW peak power and 60 fs duration, and $\gamma = 0.01 \text{ W}^{-1}\text{m}^{-1}$, $\beta_2 = -1.7 \times 10^{-27} \text{ s}^2\text{m}^{-1}$ and $\beta_3 = 6.5 \times 10^{-42} \text{ s}^3\text{m}^{-1}$.

Numerical simulations

GNLSE simulation. We model supercontinuum generation by injecting sech-type transform-limited pulses with 0.77–1.43 kW peak power and 70–130 fs duration (FWHM) ($\pm 30\%$ variation) at 830 nm center wavelength are injected into the anomalous dispersion regime of a 20 cm nonlinear fibre, including higher-order dispersion, self-steepening and Raman effect. The nonlinear coefficient of the fibre is $\gamma = 0.1 \text{ W}^{-1}\text{m}^{-1}$, and the Taylor-series expansion coefficients for dispersion at 830 nm are $\beta_2 = -5.90 \times 10^{-27} \text{ s}^2\text{m}^{-1}$, $\beta_3 = 4.21 \times 10^{-41} \text{ s}^3\text{m}^{-1}$, $\beta_4 = -1.25 \times 10^{-55} \text{ s}^4\text{m}^{-1}$, and $\beta_5 = -2.45 \times 10^{-70} \text{ s}^5\text{m}^{-1}$. The simulations use 1024 spectral/temporal grid points with temporal window size of 2 ps, and a step size of 0.02 mm (10,000 steps). For the neural network, the propagation is downsampled at a constant propagation step of $\Delta z = 0.1 \text{ cm}$, yielding 200 propagation steps. Shot noise is added via one-photon-per-mode with random phase in the frequency domain, although noise does play no significant physical role in the regime of coherent propagation studied here.

Normalized GNLSE. For generalization of the the prediction model, we used the normalized form of the GNLSE:

$$i \frac{\partial \psi}{\partial \xi} - \frac{\text{sgn}(\beta_2)}{2} \frac{\partial^2 \psi}{\partial \tau^2} - i \frac{q}{6} \frac{\partial^3 \psi}{\partial \tau^3} + \left(1 + is \frac{\partial}{\partial \tau} \right) \left(\psi \int_{-\infty}^{+\infty} r(\tau') |\psi(\tau - \tau') \xi|^2 d\tau' \right) = 0,$$

where $\tau = T/T_0$, $\xi = zT_0^2/|\beta_2|$ and $\psi(\xi, \tau) = NA(z, T)/\sqrt{P_0}$, $q = \beta_3/|\beta_2|T_0$, $s = 1/\omega_0 T_0$, and r are the normalized time, propagation distance, amplitude, third-order dispersion, shock-term, and Raman response, respectively.

For the normal dispersion, we model the propagation of transform-limited hyperbolic-secant pulses centered at 835 nm in the normally dispersive fiber (i.e. $\text{sgn}(\beta_2) = 1$). Simulations included a variation to the soliton number from 2 to 8 and dispersion parameter from 1 to 25 with the positive sign of β_2 . We used 512 temporal grid points with temporal window size of 400 (normalized units) and normalized distance set to 0.5 with 9,000 steps.

For the comparison with the recurrent neural network [15], we simulated the propagation of transform-limited hyperbolic-secant pulses centered at 830 nm in the anomalous dispersion regime (i.e. $\text{sgn}(\beta_2) = -1$). The soliton number, pulse duration and third-order dispersion parameter were randomly varied in the interval 2 to 8, 30 to 130 fs and 1 to 9, respectively. The simulations used 8192 temporal grid points with temporal window size of 350 (normalized units) and normalized distance set to 2 with 9,000 steps.